

# Accurate measurement of the H I column density from H I 21 cm absorption-emission spectroscopy

Jayaram N. Chengalur<sup>1\*</sup>, Nissim Kanekar<sup>1</sup> and Nirupam Roy<sup>2</sup>

<sup>1</sup>National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411 007, India

<sup>2</sup>Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121, Bonn, Germany

Accepted yyyy month dd. Received yyyy month dd; in original form yyyy month dd

## ABSTRACT

We present a detailed study of an estimator of the H I column density, based on a combination of H I 21 cm absorption and H I 21 cm emission spectroscopy. This “isothermal” estimate is given by  $N_{\text{HI,ISO}} = 1.823 \times 10^{18} \int [\tau_{\text{tot}} \times T_{\text{B}}] / [1 - e^{-\tau_{\text{tot}}}] dV$ , where  $\tau_{\text{tot}}$  is the total H I 21 cm optical depth along the sightline and  $T_{\text{B}}$  is the measured brightness temperature. We have used a Monte Carlo simulation to quantify the accuracy of the isothermal estimate by comparing the derived  $N_{\text{HI,ISO}}$  with the true H I column density  $N_{\text{HI}}$ . The simulation was carried out for a wide range of sightlines, including gas in different temperature phases and random locations along the path. We find that the results are statistically insensitive to the assumed gas temperature distribution and the positions of different phases along the line of sight. The median value of the ratio of the true H I column density to the isothermal estimate,  $N_{\text{HI}}/N_{\text{HI,ISO}}$ , is within a factor of 2 of unity while the 68.2% confidence intervals are within a factor of  $\approx 3$  of unity, out to high H I column densities,  $\leq 5 \times 10^{23} \text{ cm}^{-2}$  per  $1 \text{ km s}^{-1}$  channel, and high total optical depths,  $\leq 1000$ . The isothermal estimator thus provides a significantly better measure of the H I column density than other methods, within a factor of a few of the true value even at the highest columns, and should allow us to directly probe the existence of high H I column density gas in the Milky Way.

**Key words:** ISM: general – radio lines: ISM

## 1 INTRODUCTION

The neutral atomic hydrogen column density  $N_{\text{HI}}$  is an important input to our understanding of the interstellar medium (ISM). For example, it determines whether the gas is predominantly ionized (for  $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$ , in the intergalactic medium), predominantly neutral (for  $N_{\text{HI}} \gtrsim 2 \times 10^{20} \text{ cm}^{-2}$ , in typical gas clouds in galaxies), or predominantly molecular (for  $N_{\text{HI}} \gtrsim 10^{22} \text{ cm}^{-2}$ , in compact molecular clouds). It serves as the reference for estimates of various interesting quantities such as gas metallicities and abundances, is required to derive the gas spin temperature, and is the basic input for models of gas clouds. Accurate estimates of  $N_{\text{HI}}$  are thus critical for ISM studies.

There are two standard approaches towards measuring  $N_{\text{HI}}$  in Galactic clouds. The first is based on absorption spectroscopy of stars and quasars in the Lyman- $\alpha$  line, which develops wide Lorentzian wings for typical sightlines through the Milky Way (or external galaxies), and whose equivalent width is directly related to the H I column density. Such damping wings are easily detectable with modern optical spectrographs for  $N_{\text{HI}} \gtrsim 10^{19} \text{ cm}^{-2}$ , and offer accurate  $N_{\text{HI}}$  measurements for sightlines that do not contain much dust. Unfortunately, the presence of significant amounts of dust

along a sightline causes obscuration of the background star/quasar and, further, the amount of dust obscuration correlates with the total hydrogen column density in the Milky Way. As a result, it is very difficult to use Lyman- $\alpha$  spectroscopy to measure  $N_{\text{HI}}$  along high column density Galactic sightlines, with  $N_{\text{HI}} \gg 10^{21} \text{ cm}^{-2}$ .

The second approach to  $N_{\text{HI}}$  measurements is via H I 21 cm emission studies, which directly measure the brightness temperature  $T_{\text{B}}$  of the emission. For optically-thin H I 21 cm emission, the H I column density is proportional to  $T_{\text{B}}$ , even when the emission arises from multiple gas “clouds” with different temperatures. The advantage of this method is that it is easy to detect H I 21 cm emission along any Galactic sightline with today’s telescopes. Further, significant progress has recently been made in correcting for stray radiation, received through the telescope sidelobes (e.g. Kalberla et al. 2005; Bajaja et al. 2005). At low to moderate column densities,  $N_{\text{HI}} < 10^{21} \text{ cm}^{-2}$ , comparisons between  $N_{\text{HI}}$  estimates from the Lyman- $\alpha$  absorption and H I 21 cm emission approaches have typically yielded excellent agreement, to within 10% (e.g. Dickey & Lockman 1990; Wakker et al. 2011).

Unfortunately, the relation between  $T_{\text{B}}$  and  $N_{\text{HI}}$  is not straightforward for the general case of arbitrary H I 21 cm optical depth. When the optical depth is significant, one has to know both the location of different emitting components along the sightline and their individual optical depths and spin temperatures to infer  $N_{\text{HI}}$

\* E-mail: chengalu@ncra.tifr.res.in (JNC)

from the measured  $T_B$ . Assuming that the gas is optically thin only yields a lower limit on the H I column density.

Sightlines with high H I column densities are also the ones that tend to have high H I 21cm optical depths. It is thus difficult to accurately estimate  $N_{\text{HI}}$  for such sightlines using either Lyman- $\alpha$  absorption or H I 21cm emission spectroscopy. While the maximum H I column density obtained in the Leiden-Argentine-Bonn (LAB) survey (assuming optically-thin H I 21cm emission) was  $N_{\text{HI}} \approx 2 \times 10^{22} \text{ cm}^{-2}$  (Kalberla et al. 2005), significantly higher  $N_{\text{HI}}$  values have been inferred from recent modelling of H I 21cm emission data of external galaxies (Braun et al. 2009; Braun 2012) as well as Lyman- $\alpha$  absorption studies of high-redshift (and low-metallicity) gamma ray bursts and quasars (Fynbo et al. 2009; Noterdaeme et al. 2012). This raises the question of whether the gas is indeed predominantly molecular at  $N_{\text{HI}} > 10^{22} \text{ cm}^{-2}$  in the Galaxy (e.g. Schaye 2001) or whether atomic hydrogen can exist at significantly higher H I column densities. In this *Letter*, we propose a different approach to determine both the H I column density along a sightline and the error on the measurement, based on a combination of H I 21cm emission and H I 21cm absorption spectroscopy.

## 2 THE FORMALISM

For the H I 21cm line, the two observables are the H I 21cm brightness temperature  $T_B$ , measured from *emission* spectroscopy, and the H I 21cm optical depth  $\tau$ , measured from *absorption* studies towards background radio continuum sources. For a single homogeneous H I cloud, the observed brightness temperature is given by<sup>1</sup>

$$T_B = T_s \times [1 - \exp(-\tau)] , \quad (1)$$

while the H I column density  $N_{\text{HI}}$ , the H I 21cm optical depth  $\tau$  and the spin temperature  $T_s$  are related by the expression

$$N_{\text{HI}} = 1.823 \times 10^{18} \times \int T_s \tau \, dV , \quad (2)$$

where  $N_{\text{HI}}$  is in  $\text{cm}^{-2}$ , the spin and brightness temperatures are in K, and  $dV$  is in  $\text{km s}^{-1}$ , with the integral over the line profile. Note that the above equations make no approximations, except that  $T_s \gg h\nu_{21\text{cm}}/k_B \approx 0.07 \text{ K}$  (e.g. Field 1958), which should be valid in all astrophysical circumstances.

For multiple H I clouds along a sightline, equation (2) remains unchanged, except that  $T_s$  is then the column-density-weighted harmonic mean of the spin temperatures of the different clouds along the sightline; we will denote this harmonic mean spin temperature as  $\langle T_s \rangle$ . However, the expression for  $T_B$  is much more complicated in this situation (e.g. Heiles & Troland 2003a):

$$T_B = \sum_{i=0}^{N-1} T_{s,i} [1 - \exp(-\tau_i)] \times \exp\left[-\sum_{j=0}^{M_i-1} \tau_j\right] , \quad (3)$$

for  $N$  “clouds” with different spin temperatures and optical depths along the sightline, and  $M_i$  clouds between us and the  $i$ ’th cloud.

If the H I 21cm absorption is optically thin (i.e. peak optical depth  $\ll 1$ ), the above expressions can be combined to obtain

$$N_{\text{HI,OT}} = 1.823 \times 10^{18} \times \int T_B dV , \quad (4)$$

and one can estimate the H I column density directly from the H I 21cm emission spectrum. However, in the general case of arbitrary optical depth, it is not possible to determine  $N_{\text{HI}}$ , even on combining the H I 21cm absorption and emission spectra. In such a

situation, one would, in order to determine  $N_{\text{HI}}$ , need to know both the parameters  $(T_s, \tau)$  of individual clouds as well as the spatial distribution of these clouds along the sightline; the latter is especially difficult to ascertain observationally.

Our aim is to estimate the H I column density, given measurements of both the brightness temperature and the H I 21cm optical depth. For this purpose, we define a quantity  $T_{s,\text{eff}}$ , akin to the spin temperature, by the relation

$$T_B = T_{s,\text{eff}} \times [1 - \exp(-\tau_{\text{tot}})] , \quad (5)$$

where  $T_B$  is the observed brightness temperature and  $\tau_{\text{tot}}$  is the *total* optical depth. We can, without loss of generality, relate the harmonic-mean spin temperature  $\langle T_s \rangle$  along a sightline to  $T_{s,\text{eff}}$  by

$$\langle T_s \rangle = f(T_B, \tau_{\text{tot}}) \times T_{s,\text{eff}} , \quad (6)$$

where  $f(T_B, \tau_{\text{tot}})$  is some unknown function of  $T_B$  and  $\tau_{\text{tot}}$  (and which also depends on the spatial distribution of clouds along the sightline). On replacing in equation (2) for  $N_{\text{HI}}$ , we obtain

$$N_{\text{HI}} = 1.823 \times 10^{18} \int f(T_B, \tau_{\text{tot}}) \times \frac{\tau_{\text{tot}} \times T_B}{[1 - \exp(-\tau_{\text{tot}})]} dV . \quad (7)$$

Knowledge of the function  $f(T_B, \tau_{\text{tot}})$  would allow us to infer the H I column density from measurements of  $T_B$  and  $\tau_{\text{tot}}$ . Of course, this function depends on the details of the sightline. However, based on the simulations of the next section, we find that the median value of the function is  $\approx 1$ , even in the extreme cases of high  $\tau_{\text{tot}}$  and high  $N_{\text{HI}}$ , with a spread of only a factor of a few around the central value. We hence set  $f \approx 1$  to obtain

$$N_{\text{HI,ISO}} = 1.823 \times 10^{18} \int \frac{\tau_{\text{tot}} \times T_B}{[1 - \exp(-\tau_{\text{tot}})]} dV . \quad (8)$$

The above equation to estimate the H I column density was earlier proposed by Dickey & Benson (1982), but has not received much attention in the literature. Dickey & Benson (1982) refer to  $N_{\text{HI,ISO}}$  as the “isothermal” estimate of the H I column density, since equation 8 is the same as the expression for the H I column density when the H I 21cm absorption and emission arise in a single cloud with a fixed spin temperature. We will continue to use this terminology, but note that it can also be regarded as a “thin-screen” estimator as it only uses the total optical depth and is independent of the spatial disposition of the clouds. In effect, knowledge of the total H I 21cm optical depth and the total brightness temperature  $T_B$  along the sightline allow one to estimate the H I column density. Note that at large  $\tau$ , equation (8) reduces to

$$N_{\text{HI,ISO}} = 1.823 \times 10^{18} \int (\tau_{\text{tot}} \times T_B) dV . \quad (9)$$

## 3 MONTE CARLO SIMULATIONS

In this section, we use a Monte Carlo procedure to validate the “isothermal” estimator by measuring the difference between the true H I column density and the H I column densities inferred from the isothermal (and optically-thin) methods. The approach taken is the converse of the situation in the real world, where we have the measured brightness temperature and H I 21cm absorption profiles, and would like to infer the H I column density. Instead, for the purposes of the simulation, we will assume that a sightline contains some fiducial “true” H I column density  $N_{\text{HI}}$  in a narrow velocity channel, that produces both H I 21cm emission and absorption. We further assume that the H I 21cm emission and absorption in this velocity bin arise from the superposition of contributions from different temperature phases along the sightline (each, of course, with an

<sup>1</sup> Note that all quantities in this section are functions of velocity.

H I column density lower than the total H I column). We distribute the total H I column randomly between the different temperature phases, and, for each distribution, compute the observed brightness temperature and the observed total optical depth. The inferred  $T_B$  and  $\tau_{\text{tot}}$  values are then used to determine the H I column density from the isothermal and optically-thin methods, and the results compared to the known H I column density. This procedure is carried out for a large number of possible sightlines, covering a wide range of both total H I column densities and distributions of the H I column between different temperature phases.

Although we are mainly interested in the high H I column density range,  $N_{\text{HI}} \gg 10^{21} \text{ cm}^{-2}$ , the simulation has been carried out for total  $N_{\text{HI}}$  values in the range  $10^{20} \text{ cm}^{-2} \leq N_{\text{HI}} \leq 10^{24} \text{ cm}^{-2}$ . The upper end of the range was chosen so as to probe the extremely high H I columns ( $\geq 10^{23} \text{ cm}^{-2}$ ) whose presence has been suggested by modelling studies of H I in local galaxies (Braun et al. 2009; Braun 2012). The channel width is assumed to be  $1 \text{ km s}^{-1}$ , similar to the velocity resolution of the LAB all-sky H I 21cm emission survey (Kalberla et al. 2005). This narrow velocity width was chosen to match the narrowest full-width-at-half-maximum of H I 21cm lines, for purely thermal line broadening of gas at a kinetic temperature of  $\approx 20 \text{ K}$ , so that there is no loss of information due to under-sampling of the line profiles. Note that H I column densities  $\geq 10^{23} \text{ cm}^{-2}$  are extremely large for such a small velocity range; we will discuss the effect of this assumption later.

We also note that the simulation explicitly deals with the H I column density within a single *observed*  $1 \text{ km s}^{-1}$  velocity channel, instead of over some physically motivated velocity width. It is well known that H I in the ISM has velocity structure (i.e. correlations between neighbouring velocity channels), due to the gas temperature, turbulent motions, bulk motions, etc., and the net H I column density along a sightline is inferred from the full H I 21cm line profile, not merely from individual velocity channels. Our simulation makes no assumptions about the velocity structure, because quantities like the gas temperature, turbulent motions and velocity gradients are, in general, poorly known, and assumptions about these quantities could bias the results. Our approach of using narrow ( $\leq 1 \text{ km s}^{-1}$ ) channels does not lead to any loss of generality, as the channel width is narrow enough to properly sample the velocity profile. As discussed in more detail later, the primary consequence of not using the velocity structure in the line profile is that the error bars that we advocate are conservative; the advantage is that no assumptions are needed about quantities that are difficult to measure.

The next step is to specify the *spin* temperatures of the different gas clouds along the sightline. In the classic models of the ISM, H I is expected to stably exist in two phases in pressure equilibrium with each other (e.g. Field et al. 1969; Wolfire et al. 2003); these are the cold neutral medium (CNM, with  $40 \text{ K} \leq T_k \leq 200 \text{ K}$ ) and the warm neutral medium (WNM, with  $5000 \text{ K} \leq T_k \leq 8000 \text{ K}$ ). Recently, there has been evidence that significant fractions of H I may be in a thermally unstable phase, with  $200 \text{ K} \leq T_k \leq 5000 \text{ K}$  (e.g. Heiles & Troland 2003a; Kanekar et al. 2003) and that the CNM temperature range extends to  $\approx 20 \text{ K}$  (Heiles & Troland 2003b). We hence allow for gas in three temperature phases, the CNM, the WNM and the thermally unstable neutral medium (UNM), and will also use the temperature range  $20 \text{ K} \leq T_k \leq 200 \text{ K}$  for the CNM. As will be seen, the precise choice of the CNM, WNM and UNM temperature ranges do not significantly affect our results.

To estimate the brightness temperature and the H I 21cm optical depth, we require the spin temperature, not the kinetic temperature. For the CNM, the H I 21cm transition is expected to be

thermalized by collisions, causing  $T_s \approx T_k$ ; we will hence use the spin temperature range  $20 \text{ K} \leq T_s \leq 200 \text{ K}$  for the CNM. In the case of the WNM (and, possibly, the UNM), the low particle number density means that collisions cannot thermalize the line; the spin temperature is expected to typically be lower than the kinetic temperature in this phase unless there are sufficient Lyman- $\alpha$  radiation in the cloud for resonant scattering of these photons to couple the spin and kinetic temperatures (e.g. Field 1958; Liszt 2001). We have hence assumed that  $T_s < T_k$  in the WNM, and have used the results of Liszt (2001) to relate the spin temperature to the kinetic temperature; this essentially meant using the range  $2000 \text{ K} \leq T_s \leq 5000 \text{ K}$  for the WNM. The UNM spin temperature range is set to  $200 \text{ K} \leq T_s \leq 2000 \text{ K}$ , intermediate between the CNM and WNM phases. Note that this procedure effectively allows for gas at all spin temperatures  $20 \leq T_s \leq 5000 \text{ K}$ .

In order to divide the total H I column density into different components along the sightline, we first fix the fraction of gas in the three temperature phases. The simulations were carried out for the following six cases:

- (i) Half the gas is in the CNM, and half in the WNM,
- (ii) Each of the three phases contains one-third of the gas,
- (iii) 50% CNM, 40% WNM and 10% UNM,
- (iv) 50% CNM, 10% WNM and 40% UNM,
- (v) 90% CNM and 10% WNM, and
- (vi) The fraction in each phase varies randomly from run to run.

Having fixed the total H I column density and the gas fraction in each phase, we know the H I column density in each phase. We allow for the possibility that there could be multiple “clouds” from a given phase that contribute to the emission/absorption in the velocity channel by allowing gas in each phase to be further sub-divided into multiple components. We then randomly assign a spin temperature to the first component of each phase, subject to the constraint that it lies within the  $T_s$  range of the phase. The H I column density of the component is also randomly chosen from the range  $1 \times 10^{19} \text{ cm}^{-2}$  to  $5 \times 10^{22} \text{ cm}^{-2}$ . We then add new components of the same phase until the sum of their H I column densities is equal to or greater than the total H I column density set for the phase. In the latter case, the H I column density of the last component is reduced to a value which makes the sum of H I column densities match the total H I column density. In this process, if the  $N_{\text{HI}}$  of the last component falls below the minimum allowed value ( $1 \times 10^{19} \text{ cm}^{-2}$ ), all the components are discarded and the entire process is repeated. Similarly, if the brightness temperature along the line of sight exceeds  $500 \text{ K}$ , then all components are discarded and the process is repeated. Note that the highest brightness temperature observed in the Galaxy is  $\approx 200 \text{ K}$ , significantly lower than the above threshold.

The above procedure is carried out for all three phases. Once all the components in all three phases have been chosen, the order of the components along the sightline is randomized. The known H I column density and spin temperature of each component are then used to compute the two observables for the sightline, the total optical depth  $\tau_{\text{tot}}$  and the brightness temperature  $T_B$ ; these are then used in equations (4) and (8) to estimate the H I column density from the optically-thin ( $N_{\text{HI,OT}}$ ) and isothermal ( $N_{\text{HI,ISO}}$ ) methods. Of course, the true H I column density along the sightline,  $N_{\text{HI}}$ , is already known. For each bin in total “true” H I column density (bin width = 0.235 in  $\log[N_{\text{HI}}]$ ) and total H I 21cm optical depth (bin width = 0.2), we carried out 2001 runs for each of the six distributions between the different temperature phases, computing the statistics of the ratios ( $N_{\text{HI}}/N_{\text{HI,ISO}}$ ) and ( $N_{\text{HI}}/N_{\text{HI,OT}}$ ) in every case.

Finally, we also separately ran the simulations with the tem-

perature ranges  $20 \leq T_s \leq 200$  K for the CNM,  $200 \text{ K} \leq T_s \leq 5000$  K for the UNM, and  $5000 \text{ K} \leq T_s \leq 8000$  K for the WNM, i.e. assuming that the H I 21cm transition is thermalized (with  $T_s \approx T_k$ ) in all three phases. This too was done for the above six distributions of gas between the different phases. No significant difference was found between the results of the simulations using  $T_s \approx T_k$  and those using the Liszt (2001) relation between  $T_s$  and  $T_k$ ; we will hence restrict the discussion to the latter in the following sections.

#### 4 RESULTS AND DISCUSSION

In practice, qualitatively similar results were obtained for the different distributions of gas between the different temperature phases. We hence show results only for the last case, where the fraction of gas in the CNM, WNM and UNM phases is allowed to vary randomly from run to run. This is likely to provide the most robust estimate of the error on the results, unbiased by a specific choice of a gas distribution.

The critical quantity in determining the accuracy of the optically-thin and isothermal estimates of the H I column density is the total optical depth along the sightline  $\tau_{\text{tot}}$ . The accuracy of either estimate is worse at high optical depths than for  $\tau \leq 1$ , because it is easier to hide H I behind foreground gas that is optically thick. Figs. 1[A] and [B] show, respectively, the ratio of the true H I column density to the optically-thin and isothermal estimates of  $N_{\text{HI}}$ , for the case where the fractions of gas in the three temperature phases are allowed to vary randomly from run to run. As expected, the optically-thin estimate always yields a lower limit to the true H I column density, i.e. the ratio is always  $> 1$ . Even for  $\tau_{\text{tot}} \approx 1$ , the median H I column density derived from the optically-thin estimate underestimates the true H I column density by a factor of  $\approx 1.6$ ; the under-estimate is by more than an order of magnitude for optical depths of  $\approx 10$ . The spread on the optically-thin estimate is also very large, with the 68.2% confidence level on the ratio  $N_{\text{HI}}/N_{\text{HI,OT}}$  reaching  $\approx 40$  for  $\tau_{\text{tot}} \approx 10$ .

Conversely, the median isothermal estimate of the H I column density  $N_{\text{HI,ISO}}$  tracks the true H I column density to better than 10% even for optical depths of  $\approx 5$ . Even for  $\tau_{\text{tot}} \approx 10$ , the median value of  $N_{\text{HI}}/N_{\text{HI,ISO}}$  is within a factor of 1.5 of unity. The 68.2% confidence intervals on the ratio extend from  $\approx 0.4$  to  $\approx 3$ , even at high optical depths. Clearly, even for  $\tau_{\text{tot}} \approx 10$ , the isothermal estimate of the H I column density appears accurate to within a factor of  $\approx 3$ .

Figs. 2[A] and [B] show the same ratios plotted in Fig. 1, but this time as a function of the true H I column density. Note that the largest total optical depth on sightlines included in this figure is  $\approx 1000$ . The left panel [A] shows that, while the optically-thin estimate provides a lower limit to the true  $N_{\text{HI}}$ , this is not particularly useful at large values of  $N_{\text{HI}}$  ( $\gtrsim 10^{22} \text{ cm}^{-2}$ ), for which the true  $N_{\text{HI}}$  is more than an order of magnitude larger than  $N_{\text{HI,OT}}$ . On the other hand, the median value of the thin screen estimate  $N_{\text{HI,ISO}}$  tracks the true  $N_{\text{HI}}$  to within a factor of 2 over the entire H I column density range ( $N_{\text{HI}} \leq 5 \times 10^{23} \text{ cm}^{-2}$ ). Further, even the 68.2% confidence level intervals of the ratio lie within a factor of a few from unity. Clearly, the scatter in the estimated  $N_{\text{HI,ISO}}$  is quite modest even for very high H I column densities.

The simulation results thus indicate that the isothermal estimate of equation (8) provides a fairly good measure of the true H I column density, certainly good to within a factor of a few within 68.2% confidence intervals, even for high optical depths,  $\tau_{\text{tot}} \approx 1000$ , and high H I column densities,  $N_{\text{HI}} \approx 5 \times 10^{23} \text{ cm}^{-2}$ . (Note that optical depths of  $\approx 1000$  are significantly higher than

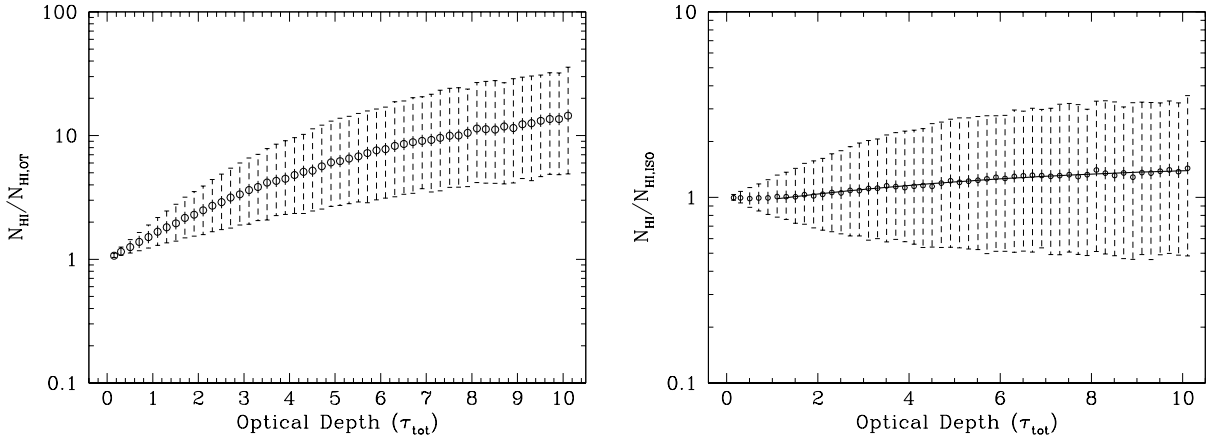
would be expected along Galactic sightlines.) This estimator should thus allow one to directly probe the existence of very high H I column densities ( $N_{\text{HI}} \gtrsim 10^{23} \text{ cm}^{-2}$ ) in the Galaxy, via accurate measurements of the H I 21cm brightness temperature and optical depth from, respectively, high velocity resolution H I 21cm emission and absorption spectroscopy. The best fit 2nd-order polynomials to the ratio of the true H I column density to the isothermal estimate of the H I column density are given in the captions to Figs. 1 and 2. Note that these are valid for  $\tau_{\text{tot}} \gtrsim 1$  and  $N_{\text{HI}} \gtrsim 10^{21} \text{ cm}^{-2}$ , respectively. Of course, at lower optical depths,  $\tau_{\text{tot}} < 1$ , the optically-thin estimate provides an acceptable measure of the true H I column density, within a factor of  $\approx 1.5$  of the true value.

It should be emphasized that the isothermal approach assumes implicitly that the H I 21cm brightness temperature and optical depth are measured *along the same sightline*. Of course, this is not the case in reality, with the optical depth along a sightline usually measured towards compact background radio sources and the brightness temperature inferred by interpolating between measurements at neighbouring positions (e.g. Heiles & Troland 2003a). Small-scale structure in the H I could imply incorrect brightness temperature estimates along the sightline and hence larger errors in the estimate of the H I column density. We do not anticipate that this will be a severe problem for most sightlines, especially given the possibility of obtaining H I 21cm emission spectra with telescopes of very different angular resolutions (e.g. from the LAB survey, the Green Bank Telescope, the Arecibo telescope, etc).

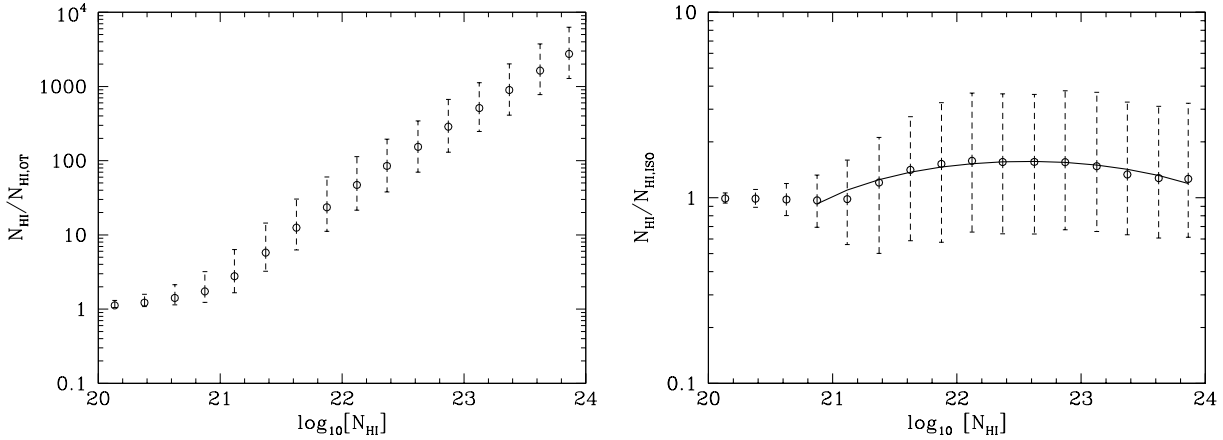
The approach also assumes that it is possible to accurately measure very high H I 21cm optical depths,  $\tau_{\text{tot}} \gg 10$ . In reality, *direct* measurements of such opacities would require spectral dynamic ranges  $\gg 10^4$  per  $1 \text{ km s}^{-1}$  channel, that will not be achieved even with future facilities like the Square Kilometer Array. However, it should be possible (except for cases with severe line blending) to determine the opacity at line peak via a Gaussian-fitting procedure to spectra of a high velocity resolution, as the wings of the features (which have lower optical depths) would also have been measured at very high signal-to-noise ratios. Note that the highest H I 21cm opacities would arise for cold gas with high H I column density, which is likely to be thermalized and, hence, to have a Gaussian line profile. We have verified from simulations that opacities of  $\gtrsim 50$  can be easily measured via such a Gaussian-fitting procedure, from spectra with velocity resolutions of  $\approx 0.5 \text{ km s}^{-1}$  and optical depth RMS noise values of  $\approx 0.001$  per  $0.5 \text{ km s}^{-1}$  channel (which have already been achieved with today's interferometers; e.g. Braun & Kanekar 2005).

The present simulations have been carried out so as to match observations with high velocity resolution,  $\lesssim 1 \text{ km s}^{-1}$ . This is to ensure that there is no loss of information due to smoothing of narrow absorption from cold H I clouds, even at  $T_k \approx 20$  K. The quoted errors above on the inferred H I column density are per  $1 \text{ km s}^{-1}$  spectral channel. However, as noted earlier, the velocity width of absorption/emission from individual gas “clouds” with  $T_k \gg 20$  K will always be significantly larger than  $1 \text{ km s}^{-1}$ . As a result, the error introduced when estimating the total H I column density along a sightline by integrating the per channel isothermal estimate across the line profile will be correlated between groups of neighbouring  $1 \text{ km s}^{-1}$  channels. A conservative estimate of the net systematic error on the total H I column density along the sightline can be obtained by assuming that the errors are perfectly correlated over the line profile, implying that the error on the total H I column density is the same as that one on the H I column density per  $1 \text{ km s}^{-1}$  velocity channel, as determined from the present simulation.

We also note that the assumption of H I column densities of



**Figure 1.** The ratio of the true H I column density  $N_{\text{HI}}$  to the H I column density obtained from [A] the optically-thin estimate  $N_{\text{HI,OT}}$  (left panel) and [B] the isothermal estimate  $N_{\text{HI,ISO}}$  (right panel), plotted against total optical depth  $\tau_{\text{tot}}$  (with a bin width of 0.2). In both panels, the filled circles mark the median value of the ratio, while the error bars encompass the 68.2% confidence intervals. For each bin, the median and the confidence interval were computed over a total of 2001 runs. The simulation is for a random distribution of gas between the three temperature phases, using the Liszt (2001) relation between spin and kinetic temperatures. The solid line in the right panel shows the best 2nd-order polynomial fit to the relation between  $[N_{\text{HI}}/N_{\text{HI,ISO}}]$  and  $\tau_{\text{tot}}$ , valid for  $\tau_{\text{tot}} > 1$ :  $[N_{\text{HI}}/N_{\text{HI,ISO}}] = 0.904 + 0.074\tau_{\text{tot}} - 0.0026\tau_{\text{tot}}^2$ .



**Figure 2.** The ratio of the true H I column density  $N_{\text{HI}}$  to [A] the optically-thin estimate  $N_{\text{HI,OT}}$  (left panel) and [B] the isothermal estimate  $N_{\text{HI,ISO}}$  (right panel), plotted against  $\log[N_{\text{HI}}]$  (with a bin width of 0.235 in  $\log[N_{\text{HI}}]$ ). In both panels, the filled circles mark the median value, while the error bars give the 68.2% confidence intervals. For each bin, the median and the confidence interval were computed from 2001 runs. The simulation is for a random distribution of gas between the three temperature phases, using the Liszt (2001) relation between spin and kinetic temperatures. The solid line in the right panel shows the best 2nd-order polynomial fit to the relation between  $[N_{\text{HI}}/N_{\text{HI,ISO}}]$  and  $\log[N_{\text{HI}}]$ , valid for  $N_{\text{HI}} \geq 10^{21} \text{ cm}^{-2}$ :  $[N_{\text{HI}}/N_{\text{HI,ISO}}] = -112.88 + 10.144 \times \log[N_{\text{HI}}/\text{cm}^{-2}] - 0.2248 \times \log[N_{\text{HI}}/\text{cm}^{-2}]^2$ .

$\approx 10^{23} - 10^{24} \text{ cm}^{-2}$  per  $1 \text{ km s}^{-1}$  velocity channel is quite extreme. Extremely high H I column densities,  $\geq 10^{23} \text{ cm}^{-2}$ , are more likely to be distributed over a significantly larger velocity range, with far smaller columns arising per  $1 \text{ km s}^{-1}$  channel. Allowing for such large column densities per  $1 \text{ km s}^{-1}$  channel results in both shifting the median value of the ratio  $N_{\text{HI}}/N_{\text{HI,ISO}}$  slightly away from unity as well as significantly increasing the error bars in the isothermal estimate. For example, if we limit to  $N_{\text{HI}} = 10^{22} \text{ cm}^{-2}$  per  $1 \text{ km s}^{-1}$  velocity channel, the median  $N_{\text{HI}}/N_{\text{HI,ISO}}$  is  $\leq 1.2$  for  $\tau_{\text{tot}} < 10$ , while the 68.2% confidence level intervals on the ratio extend from  $\approx 0.5$  to  $\approx 2$ . Our choice of  $10^{24} \text{ cm}^{-2}$  as the limiting H I column density per  $1 \text{ km s}^{-1}$  channel is thus a very conservative one.

Finally, other estimators of the H I column density have been used in the literature (e.g. Lockman & Savage 1995; Wakker et al.

2011), albeit without a detailed characterization of their accuracy in the high  $N_{\text{HI}}$  regime. For example, Wakker et al. (2011) use an estimator that appears similar to the isothermal estimate, with

$$N_{\text{HI}} = 1.823 \times 10^{18} \times \int T_s \log \left[ \frac{T_s}{T_s - T_B} \right] dV, \quad (10)$$

and assume  $T_s = 135 \text{ K}$  for the half of the spectrum with the highest  $T_B$  values and  $T_s = 5000 \text{ K}$  for the half of the spectrum with the lowest  $T_B$  values. While this has the advantage that only the H I 21cm emission spectra are needed to determine the H I 21cm column density (as is also true for the optically-thin estimate), it is not very useful at high optical depths, where  $T_s \approx T_B$ . It is in the high opacity regime that the isothermal estimate is clearly superior to the other methods.

## 5 SUMMARY

We study in detail the “isothermal” estimator of the H I column density, earlier proposed by Dickey & Benson (1982), based on measurements of the total H I 21cm brightness temperature  $T_B$  and the total H I 21cm optical depth  $\tau_{\text{tot}}$  from high velocity resolution spectroscopy. The “isothermal” estimate is given by  $N_{\text{HI,ISO}} = 1.823 \times 10^{18} \int [\tau_{\text{tot}} \times T_B] / [1 - e^{-\tau_{\text{tot}}}] dV$ . We have carried out a Monte Carlo simulation of realistic sightlines, including gas in different phases and random locations along the path, to determine the accuracy of the isothermal estimate of the H I column density. We find that this approach yields accurate estimates of the H I column density, and that the results do not strongly depend (statistically) on the spatial distribution of gas along the sightline, or its distribution in different temperature phases. In general, the median isothermal estimate of the H I column density is within a factor of 2 of the true H I column density, while the 68.2% confidence intervals of the ratio of  $N_{\text{HI}}/N_{\text{HI,ISO}}$  are within a factor of  $\approx 3$  of unity, out to extremely high H I column densities,  $\leq 5 \times 10^{23} \text{ cm}^{-2}$  per  $1 \text{ km s}^{-1}$  channel, and high total optical depths,  $\leq 1000$ . The 68.2% confidence intervals are conservative, as they allow for the possibility of extremely high H I column densities,  $> 10^{23} \text{ cm}^{-2}$ , in a narrow velocity range ( $1 \text{ km s}^{-1}$ ), which is unlikely to arise in reality. We conclude that the isothermal estimator allows one to accurately measure the H I column density (to within a factor of a few even at the highest H I column densities) and to thus directly probe the existence of extremely high H I column density gas in the Galaxy.

## ACKNOWLEDGEMENTS

NK acknowledges support from the Department of Science and Technology through a Ramanujan Fellowship. NR acknowledges the Jansky Fellowship Program of NRAO/NSF/AUI and support from the Alexander von Humboldt Foundation. We thank Robert Braun and an anonymous referee for comments on an earlier version of this paper, and Harvey Liszt for providing us with a table giving the relation between spin and kinetic temperatures in his model.

## REFERENCES

- Bajaja E., Arnal E. M., Larrarte J. J., Morras R., Pöppel W. G. L., Kalberla P. M. W., 2005, *A&A*, 440, 767  
 Braun R., 2012, *ApJ*, 749, 87  
 Braun R., Kanekar N., 2005, *A&A*, 436, L53  
 Braun R., Thilker D. A., Walterbos R. A. M., Corbelli E., 2009, *ApJ*, 695, 937  
 Dickey J. M., Benson J. M., 1982, *AJ*, 87, 278  
 Dickey J. M., Lockman F. J., 1990, *ARA&A*, 28, 215  
 Field G. B., 1958, *Proc. I. R. E.*, 46, 240  
 Field G. B., Goldsmith D. W., Habing H. J., 1969, *ApJ*, 155, L149  
 Fynbo J. P. U. et al. 2009, *ApJS*, 185, 526  
 Heiles C., Troland T. H., 2003a, *ApJS*, 145, 329  
 Heiles C., Troland T. H., 2003b, *ApJ*, 586, 1067  
 Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, *A&A*, 440, 775  
 Kanekar N., Subrahmanyam R., Chengalur J. N., Safouris V., 2003, *MNRAS*, 346, L57  
 Liszt H., 2001, *A&A*, 371, 698  
 Lockman F. J., Savage B. D., 1995, *ApJS*, 97, 1  
 Noterdaeme P. et al. 2012, *A&A*, 547, L1

Schaye J., 2001, *ApJ*, 562, L95

Wakker B. P., Lockman F. J., Brown J. M., 2011, *ApJ*, 728, 159

Wolfire M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, *ApJ*, 587, 278